

MT09 Constrained Training Condition Resources

April 02, 2009 Modifications:

replaced: LDC2006T02 Arabic Gigaword Second Edition
with: LDC2007T40 Arabic Gigaword Third Edition

replaced: LDC2005T14 Chinese Gigaword Second Edition
with: LDC2007T38 Chinese Gigaword Third Edition

replaced: LDC2005T12 English Gigaword Second Edition
with: LDC2007T07 English Gigaword Third Edition

replaced: LDC2005T01 Chinese Treebank 5.0
with: LDC2007T36 Chinese Treebank 6.0

renamed: LDC2006E26 GALE Financial News
to: LDC2006E26 GALE Y1 – English Chinese Parallel Financial News

LDC2004T18 Arabic English Parallel News Part 1
LDC2007T40 Arabic Gigaword Third Edition
LDC2004T17 Arabic News Translation Text Part 1
LDC2005E46 Arabic Treebank English Translation
LDC2005T02 Arabic Treebank: Part 1 v 3.0 (POS with full vocalization + syntactic analysis)
LDC2004T02 Arabic Treebank: Part 2 v 2.0
LDC2005T20 Arabic Treebank: Part 3 (full corpus) v2.0 (MPG + Syntactic Analysis)
LDC2004L02 Buckwalter Arabic Morphological Ana
LDC2005T34 Chinese <-> English Name Entity Lists (v1.0)
LDC2005T10 Chinese English News Magazine Parallel Text
LDC2007T38 Chinese Gigaword Third Edition
LDC2005T06 Chinese News Translation Text Part 1
LDC2005T23 Chinese Proposition Bank 1.0
LDC2007T36 Chinese Treebank 6.0
LDC2003E07 Chinese Treebank English Parallel Corpus
LDC2002L27 Chinese-English Translation Lexicon Version 3.0
LDC2007T07 English Gigaword Third Edition
LDC2004E72 eTIRR Arabic English News Text
LDC2003E14 FBIS Multilanguage Texts
LDC2004T08 Hong Kong Parallel Text
LDC2006T04 Multiple Translation Chinese (MTC) Part 4
LDC2003T18 Multiple-Translation Arabic (MTA) Part 1
LDC2005T05 Multiple-Translation Arabic (MTA) Part 2
LDC2003T17 Multiple-Translation Chinese (MTC) Part 2
LDC2004T07 Multiple-Translation Chinese (MTC) Part 3
LDC2002T01 Multiple-Translation Chinese Corpus
LDC2006E44 TIDES MT 2004 Arabic evaluation data
LDC2006E43 TIDES MT 2004 Chinese evaluation data
LDC2006E39 TIDES MT 2005 Arabic evaluation data
LDC2006E38 TIDES MT 2005 Chinese evaluation data
LDC2004E13 UN Arabic English Parallel Text
LDC2004E12 UN Chinese English Parallel Text
LDC2002E18 Xinhua Chinese English Parallel News Text Version 1 beta
LDC2000T50 Hong Kong Hansards Parallel Text

| | |
|-------------|---|
| LDC2000T47 | Hong Kong Laws Parallel Text |
| LDC2000T46 | Hong Kong News Parallel Text |
| LDC2007T08 | ISI Arabic-English Automatically Extracted Parallel Text |
| LDC2007T09 | ISI Chinese-English Automatically Extracted Parallel Text |
| LDC2007T07 | English Gigaword Third Edition |
| LDC2007T21 | OntoNotes v 1.0 |
| LDC2007T03 | Tagged Chinese Gigaword |
| LDC2003T07 | Arabic Treebank: Part 1 - 10K-word English Translation |
| LDC2007T02 | English Chinese Translation Treebank v 1.0 |
| LDC2006T10 | English-Arabic Treebank v 1.0 |
| LDC2006T13 | Web 1T 5-gram Version 1 |
| LDC2005E83 | GALE Y1 Q1 Release - Translations |
| LDC2006E85 | GALE Y1 Q3 Release - Translations |
| LDC2006E92 | GALE Y1 Q4 Release - Translations |
| LDC2006E24 | GALE Y1 - Interim Release: Translations |
| LDC2006E34 | GALE Y1 Q2 Release - Translations V2.0 |
| LDC2007E59 | MT06 Evaluation Source Text and Reference Translations |
| LDC99T42 | Treebank-3 |
| LDC2006E26 | GALE Y1 – English Chinese Parallel Financial News |
| LDC2002T31 | AQUAINT |
| LDC2005T28 | HARD |
| LDC2005T35 | ANC |
| LDC93T1 | ACL/DCI |
| LDC93T3A | TIPSTER |
| LDC95T21 | NAN |
| LDC95T6 | CSR-III |
| LDC98T30 | NAN Supplement |
| LDC2005E85 | GALE Y1 Q1 Release - English Translation Treebank |
| LDC2006E25 | GALE Y1 - Arabic English Parallel News Text |
| LDC2006E36 | GALE Y1 Q2 Release - English Translation Treebank |
| LDC2006E82 | GALE Y1 Q3 Release - English Translation Treebank |
| LDC2006E95 | GALE Y1 Q4 Release - English Translation Treebank |
| LDC2006E86 | GALE Y1 Q3 Release - Word Alignment |
| LDC2006E93 | GALE Y1 Q4 Release - Word Alignment |
| LDC2007E111 | LCTL Urdu Morphological Analyzer v2.0 |
| LDC2009E08 | NIST Open MT 2008 Current Test Set Arabic-to-English |
| LDC2009E09 | NIST Open MT 2008 Current Test Set Chinese-to-English |
| LDC2009E10 | NIST Open MT 2008 Current Test Set English-to-Chinese |
| LDC2009E11 | NIST Open MT 2008 Current Test Set Urdu-to-English |
| LDC2009E12 | NIST Open MT08 Urdu Resources |